A Theory of
Modeling Correlations for Use in Cost-Risk Analysis

Stephen A. Book

Chief Technical Officer
MCR, LLC
390 No. Sepulveda Blvd.
El Segundo, CA 90254
(310) 640-0005 x244
sbook@mcri.com

Third Annual Project Management Conference
National Aeronautics and Space Administration
Galveston TX
21-22 March 2006

1

Acknowledgment

The presenter would like to acknowledge the important role of

Dr. Patrick L. Smith
Principal Director
Architecture and Design Subdivision
The Aerospace Corporation
El Segundo CA

in the origination and development of the ideas on which this presentation is based. Dr. Smith, in fact, came up with the original concept of linking probable costs of different WBS elements through their shared characteristic of "newness." The rest of the presentation is simply mathematics. It is only by Dr. Smith's vigorous insistence that he is not listed as a co-author.

Contents

- **Cost-Risk Analysis**
- **Limits of Engineering Judgment**
- **The Trouble with Inputs**
- **A Quantification Model**
- **Modeling Inter-Element Correlation**
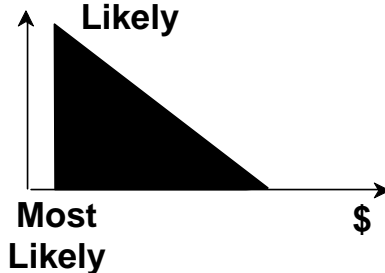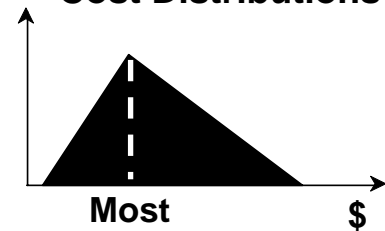- **Examples**
- **Summary**

Contents

- **Cost-Risk Analysis**
- Limits of Engineering Judgment
- The Trouble with Inputs
- A Quantification Model
- Modeling Inter-Element Correlation
- Examples
- Summary

- **"Cost Risk": A Working Definition**
  - Inadequacy of Forecasted Funding Requirements to Assure That Program Can Be Completed and Meet Its Stated Objectives
- **"Cost-Risk Analysis": A Procedure**
  - Model WBS-element Costs As Uncertain Quantities (i.e., Random Variables) That Have Probability Distributions
  - Sum WBS-element Cost Distributions Statistically (e.g., By Monte Carlo Sampling) to Generate Cumulative Distribution of Total Program Cost
  - Quantify Confidence in "Point" Estimate of Program Cost or in Whatever Amount May Have Been Budgeted for the Program
  - Read off 70th Percentile Cost, 90th Percentile Cost, etc., From Cumulative Distribution to Estimate Additional Amount of Dollars Needed to Cover Risk
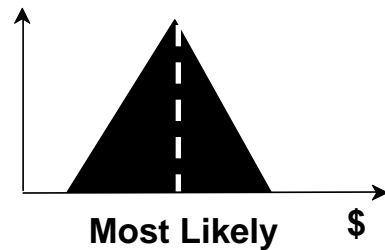
When WBS Elements Are Many
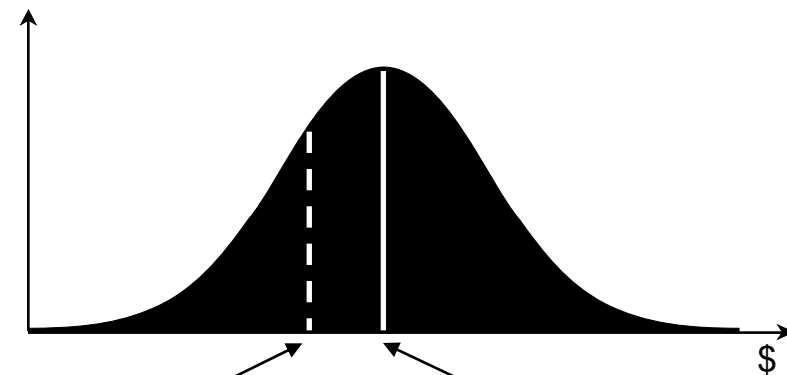
WBS-Element Triangular Cost Distributions

Most Likely — $

Most Likely — $

Most Likely — $

Merge WBS-Element Cost Distributions Into Total-Cost Normal Distribution
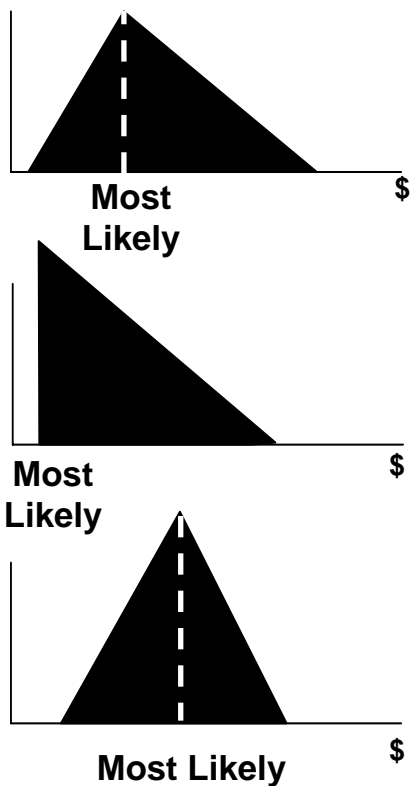
$

Roll-up of Most Likely WBS-Element Costs

Most Likely Total Cost

When WBS Elements Are Few

WBS-Element Triangular Cost Distributions

Merge WBS-Element Cost Distributions Into Total-Cost Lognormal Distribution
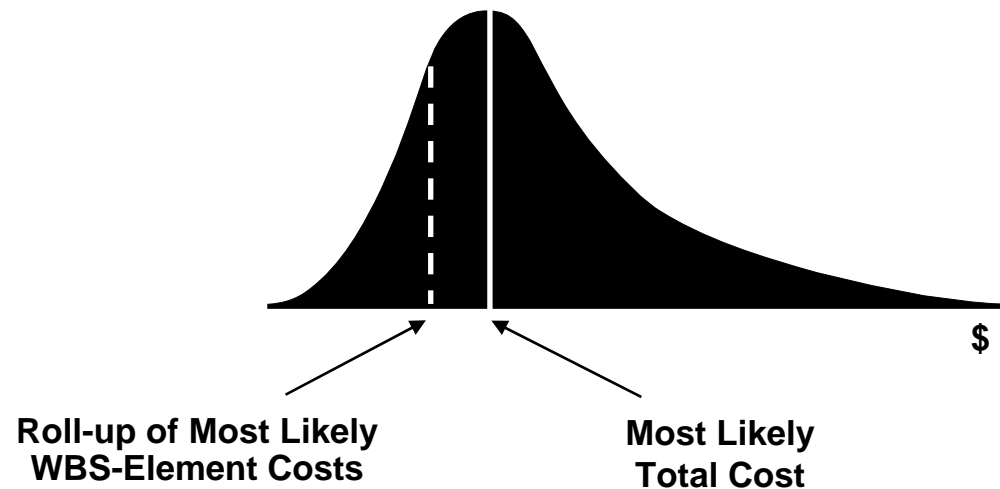
Most Likely

$

Most Likely

$

Most Likely

$

Roll-up of Most Likely WBS-Element Costs

Most Likely Total Cost

$

THE AEROSPACE CORPORATION

Contents

- **Cost-Risk Analysis**
- **Limits of Engineering Judgment**
- **The Trouble with Inputs**
- **A Quantification Model**
- **Modeling Inter-Element Correlation**
- **Examples**
- **Summary**

Ground Rules for Proposed Theory

- **WBS-Element Technical Descriptions Known to Fidelity Level Appropriate for Project Phase, Usually from CARD**
- **Traditional Cost Estimate Already Completed, Providing "Point Estimate" of Each WBS-Element Cost**
- **"Point Estimate" Assumed to be Located at Center of Normal Distribution that Models "Random" Estimating Error**
- **Risk-Driver Cost-Impact Study Provides Estimate of Highest Possible (Worst-case) Cost of Each Element, Ideally Based on Information in Risk-Management Plan**

Inter-Element Correlations are Not Zero

- **Technical, Schedule, Other Risks Impact Costs of Several Elements of the Program's WBS**
- **Some Circumstances Lead to Positive Correlations**
  - Expensive Technical Fix in One Area May Require Costly Technical Adjustments in Other Areas
  - Schedule Slip Experienced While Awaiting Solution in One Area Forces "Standing Army" Expenditures in Other Areas
- **Other Circumstances Lead to Negative Correlations**
  - Risk Dollars Judiciously Applied in One Area May Reduce Costs in Other Areas
  - Technical Fix in One Area May Be Applicable to (and Lessen Risk Without Much Additional Expense in) Other Areas
- **Numerical Values of these Correlations are Difficult to Establish, but Nevertheless Estimates of Them Must be Made**

Correlation Matters

Percent by which Total-Cost Sigma is Underestimated When Correlation Assumed to be 0 Instead of ρ
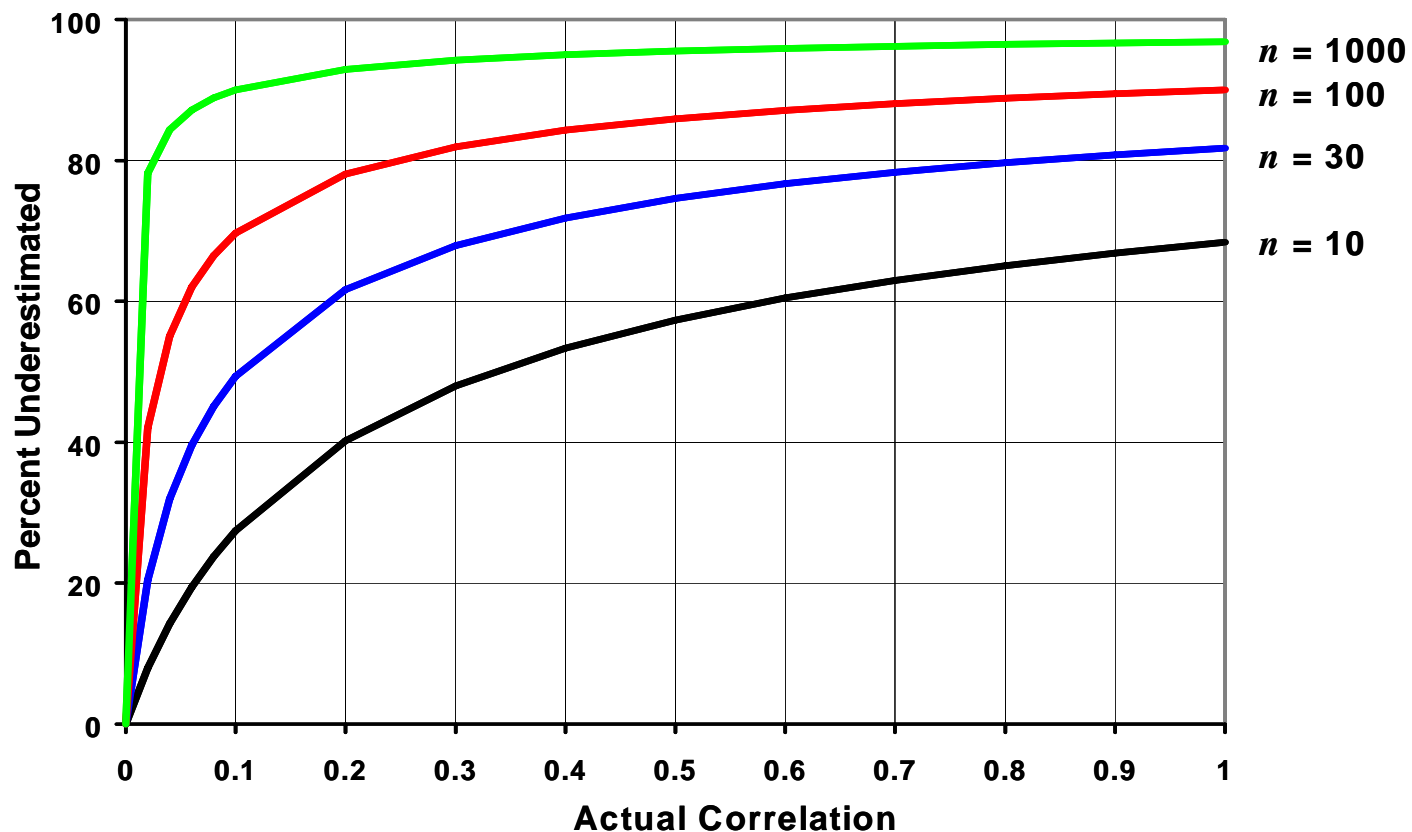
Note: A Generic Chart Based on WBS in which All n Elements Have Equal Sigma Values

- "Ignoring" Correlation Issue is Equivalent to Assuming that Risks are Uncorrelated, i.e., that All Correlations are Zero
- Reasonable Choice of Nonzero Values Brings You Closer to Truth
  - Most Elements are, in Fact, Pairwise Correlated
  - 0.2 is at "Knee" of Curve on Previous Chart, thereby Providing Most of the Benefits at Least Commitment
- Square of Correlation (namely, R^2) Represents Percentage of Variation in one WBS Element's Cost that is Attributable to Influence of Another's

Correlation	% Influenced
0.00	0%
±0.10	1%
±0.32	10%
±0.50	25%
±0.71	50%

12

Contents

- **Cost-Risk Analysis**
- **Limits of Engineering Judgment**
- **The Trouble with Inputs**
- **A Quantification Model**
- **Modeling Inter-Element Correlation**
- **Examples**
- **Summary**

Inputs to Cost-Risk Models

- **Several Cost-Risk Software Packages (e.g., FRISK, @Risk™, Crystal Ball™, Risk+) Accept a Full Correlation Matrix as Input**
- **This Means They Need Many Input Parameters in Order to Work at Maximum Capability**

Number of WBS Elements*	Added Parameters for Correlation	Number of Input Values Required
5	10	25
10	45	75
15	105	150
20	190	250
50	1225	1375
n	$n(n-1)/2$	$(n^2 + 5n)/2$

*Costs of which are modeled as triangular distributions.

- **We Are Not Free to Assign Correlation Values Arbitrarily**
  - Correlations Must be Consistent among Themselves
  - To Ensure Consistency, Range of Allowable Numerical Values of Correlation Coefficients is Subject to Strict Mathematical Constraints
  - Correlation "Matrix" Must Be Nonnegative Definite, i.e., Have No Negative Eigenvalues
- **We Need a Procedure that will Allow Us to Avoid Getting Tripped Up by the Mathematical Constraints and Yet Model the Real Situation in Any Particular Case as Well as Possible**

- … **To Generate the Large Number of Barely Understood Parameters**
- … **To Apply Cost-Analysts' and Engineers' Judgments of Risk Issues**
- … **Support Cost-Risk Software Packages' Need for Correlation Matrices**
- … **To Ensure that the Correlation Matrix is Self-Consistent**

Contents

- **Cost-Risk Analysis**
- **Limits of Engineering Judgment**
- **The Trouble with Inputs**
- **A Quantification Model**
- **Modeling Inter-Element Correlation**
- **Examples**
- **Summary**

- **Assuming Zero Rather than Positive Correlation Values is Too Optimistic, because Summing Uncorrelated Random Variables Exerts Cancellation Effect on Total-Cost Standard Deviation ("sigma value")**
- **Consider a Model Based on a Theory of How Correlations Arise between WBS Elements**
  - Correlations Originate in Weight (and Resulting Cost) Growth As System Undergoes Development
  - Problems Encountered in One Subsystem Often Induce Weight and Cost Increase in Associated Subsystems
- **Historical Observation: A Large Proportion of Cost Growth in a Program Can Often be Traced to a Single Common Source or a Small Group of Related Sources**

Model Outline

- **Inputs**
  - "Point Estimate" of Cost
  - Standard Error of Estimate (Estimating Error)
  - Percentage of Technology, Design, Software, etc., that Is New
  - To-be-Explained-Later "Cost-growth Sensitivity Factor"
- **Mathematics: Statistical Summation**
  - Point Estimate
  - Normal Error Component (Defined by Standard Error of Estimate)
  - Risk Component Assumed to Have Triangular Distribution (Defined by Cost-growth Sensitivity Factor)
- **Outputs**
  - Correlations Among Element Cost Variations
  - Mathematically Valid Correlation-Matrix Inputs to Risk-Analysis Software

B

Point-Estimate WBS-Element Cost

0

Distribution of WBS-Element Estimating Error

DENSITY

L-B   0   **O**   H-B $

Optimistic Cost

Best-Estimate Cost (Mode)

Cost Implication of Technical Assessment

B

Distribution of WBS-Element Cost

20
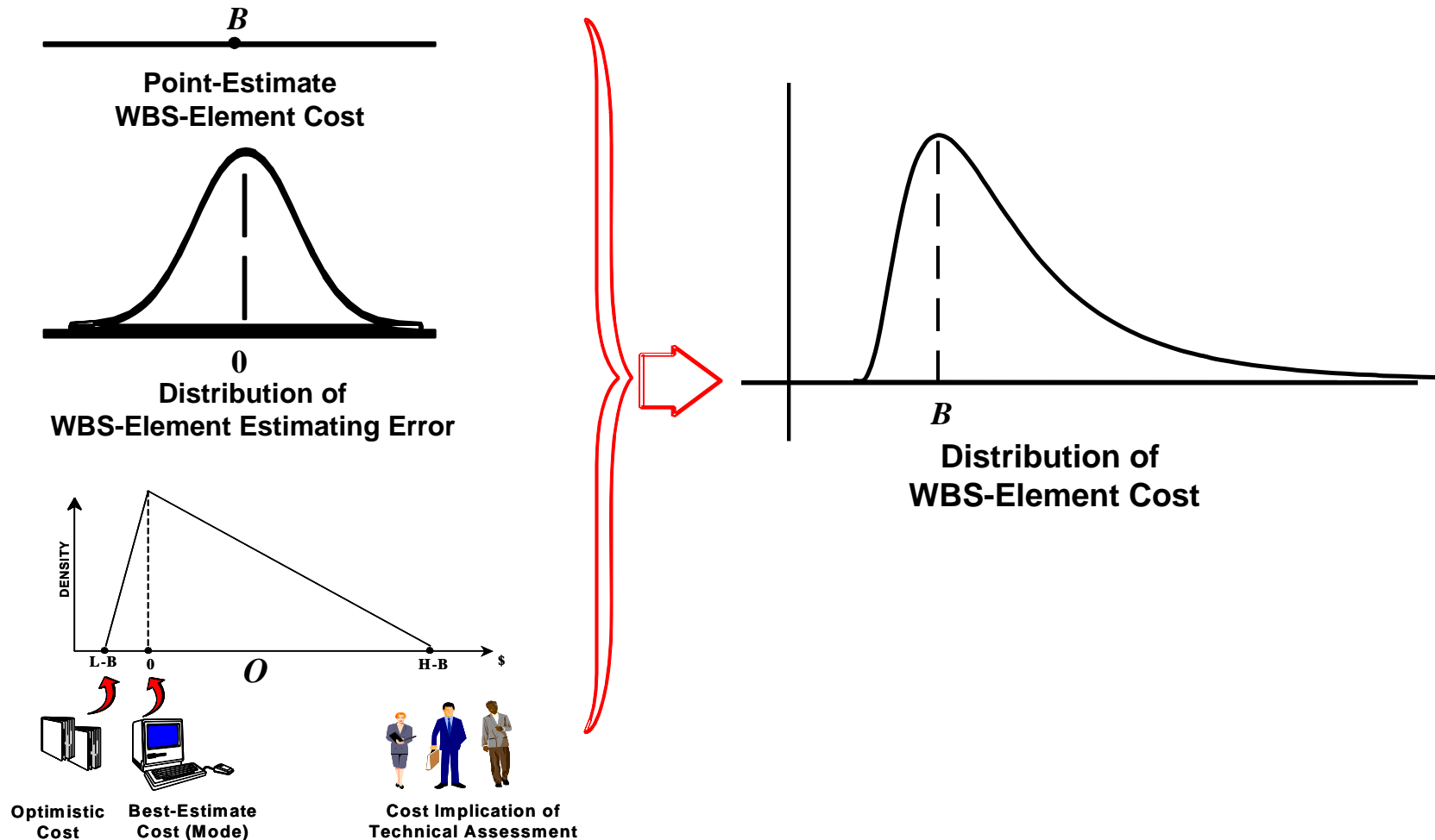
Ideal Cost-Risk Procedure

DISTRIBUTIONS OF
WBS-ELEMENT COSTS

Space

B_1

Launch

B_2

Ground
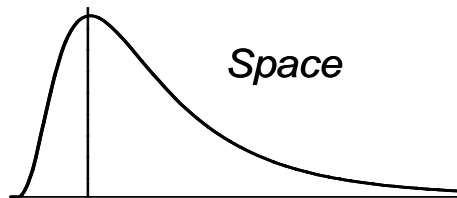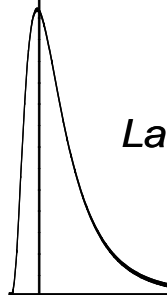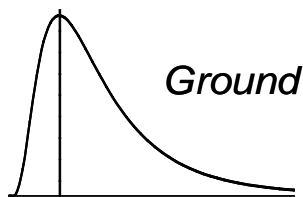
B_3

$$\begin{bmatrix} 1 & 0.5 & 0 \\ 0.5 & 1 & .3 \\ 0 & .3 & 1 \end{bmatrix}$$

Correlation Matrix

MERGE WBS-ELEMENT COST
DISTRIBUTIONS AND CORRELATIONS
INTO TOTAL-COST DISTRIBUTION

70% PROBABILITY

$

Point ESTIMATE COST
(MOST LIKELY)

RISK
DOLLARS

70th PERCENTILE
COST
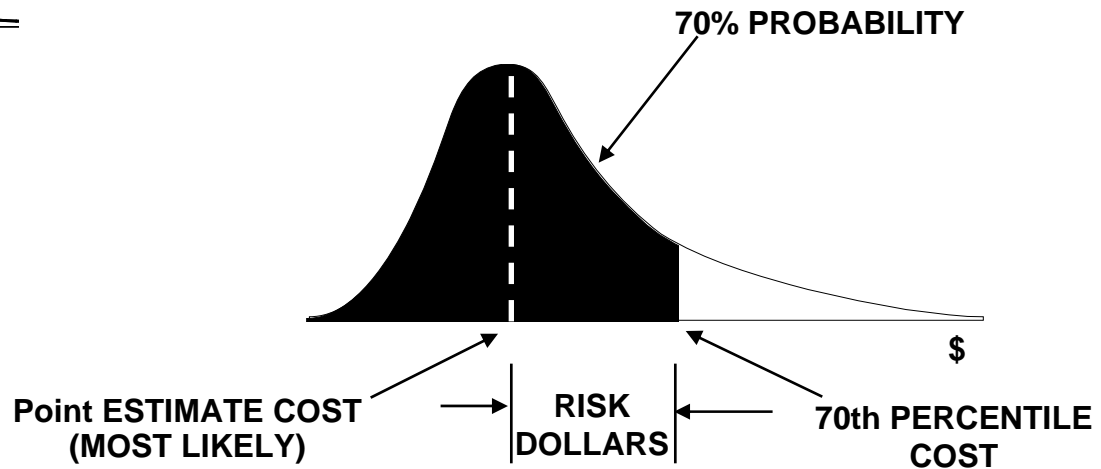
Note: Addition of risk dollars brings confidence that
total appropriation (Point estimate plus risk dollars)
is sufficient to fund program.

21

Contents

- **Cost-Risk Analysis**
- **Limits of Engineering Judgment**
- **The Trouble with Inputs**
- **A Quantification Model**
- **Modeling Inter-Element Correlation**
- **Examples**
- **Summary**

Ground Rules for Estimating Inter-Element Correlations

- **Apply Usual Methods to Establish Point Estimate of Cost for Each WBS Element**
- **Model Cost-Estimating Uncertainties as Random Variables**
  - **Normally Distributed With Zero Mean**
  - **Standard Deviations are those Associated With Estimating Method for Each Individual WBS Element**
  - **Pairwise Correlations are Zero**
- **Model Cost-Growth Impacts of Risk Issues as Random Variables**
  - **Triangular Distributions**
  - **Assume Maximum Possible (Worst-case) Cost Proportional to Specified Percentage of Element Cost Especially Subject to Growth (Percent "New")**
  - **Assume Cost Growth "Newness" Portions are Correlated 100% (or some other percentge), e.g., Have Common Cause, among All WBS Elements**

23

- B_i =   Point-estimate Cost of Element i
- E_i =   Normal $N(0, U_i)$ Random Variable, Representing Estimating Error of Cost of Element i
- P_i =   "Newness" Fraction of Element i (by weight, function, or other relevant measure)
- k =   **Cost-Growth Sensitivity Factor (more about this later)**
- R_i =   Right-Triangle $T(0, 0, k\, P_i\, B_i)$ Random Variable, Representing Technical Risk of Element i Cost
- $Corr\left(R_i, R_j\right)$ = ρ (= 1.00 for this presentation)
- C_i = Cost of Element i = $B_i + E_i + R_i$
- $\rho_{ij} = Corr\left(C_i, C_j\right) = \dfrac{Cov\left(R_i, R_j\right)}{\sqrt{Var\left(C_i\right)\, Var\left(C_j\right)}}$

24

Cost Estimating Uncertainty E_i
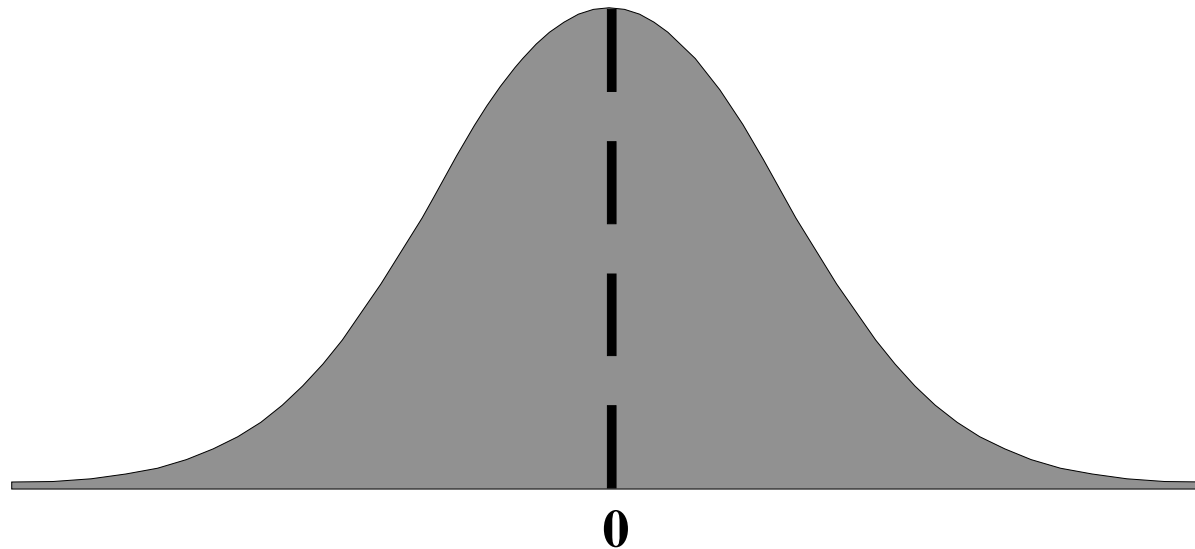
0 = Mean of Cost-Estimate Uncertainty

U_i = Standard Deviation of Cost-Estimate Uncertainty
  (Sigma Value of Normal Error Distribution)

Military Space Percentage Weight Growth (As Function of New Technology Percentage)

k = Cost-Growth Sensitivity Factor (Worst-Case Measure of Extent to which "Newness" Fraction is Expected to Impact Cost Growth)

$k = 0.6$
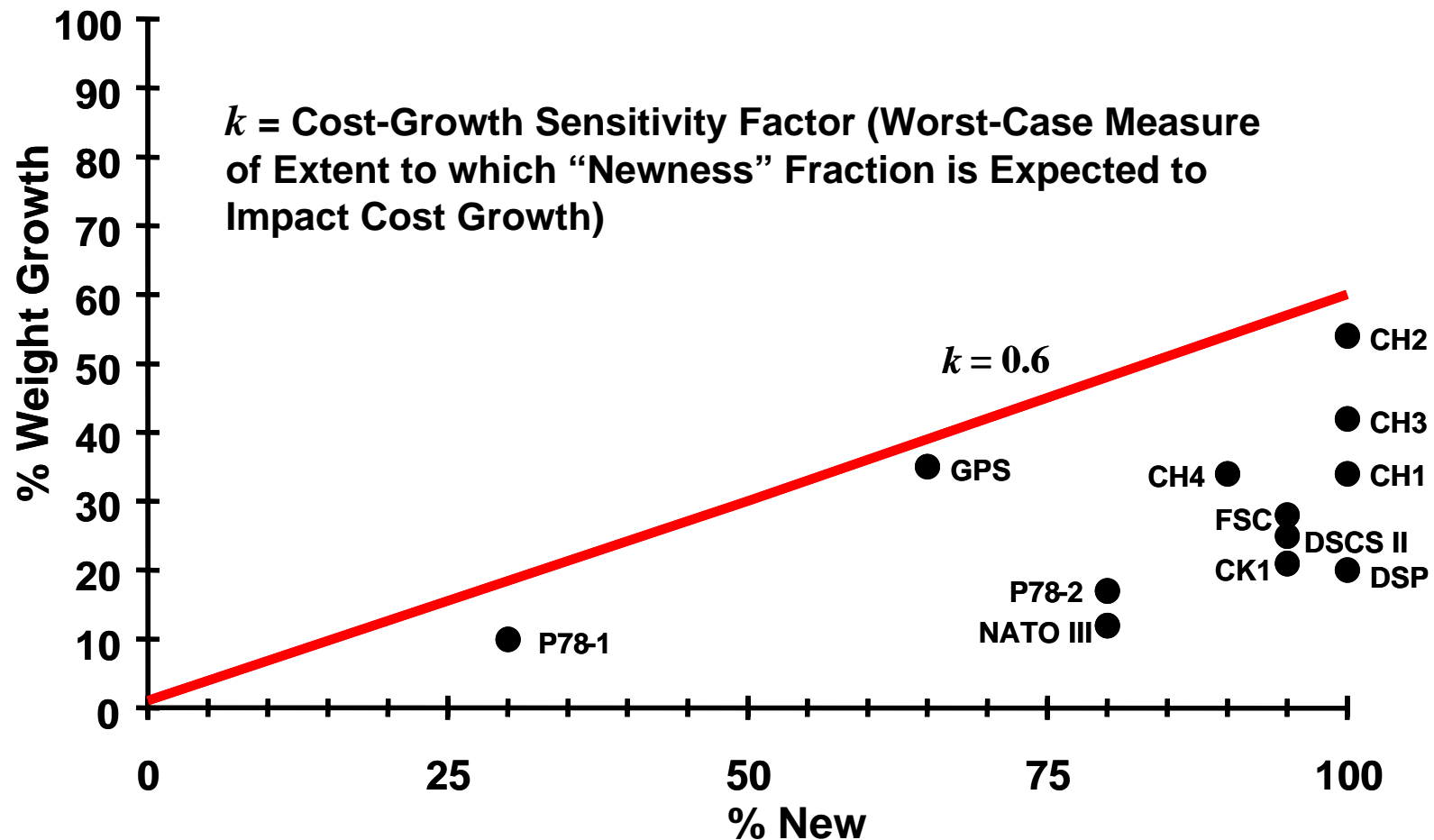
k = Upper Bound on Ratio: Weight Growth to New Technology Percentages

THE AEROSPACE CORPORATION

A Short Bibliography of Studies on Cost Growth

- R.L. Coleman, J. Summerville, and M. Dameron, "Schedule and Cost Growth," 35th DoD Cost Analysis Symposium (February 2002), SCEA National Conference (June 2002), PMI 2002 National Conference (November 2002), 62 charts.
- C. Cummings, M. Gallo, P. Johnson, B. Marsh-Jones, and J. von Kuegelgen,, *Software Development Estimating Handbook, Phase One*, Naval Center for Cost Analysis, February 1998, 164 pages. (http://www.ncca.navy.mil)
- J. Drezner, J. Jarvaise, R. Hess, P. Hough, and D. Norton, *An Analysis of Weapon System Cost Growth*, The RAND Corporation, Document No: MR-291-AF, 1993, 84+xix pages.
- C. Edwards, "Government schemes cost more than promised," *Cato Institute Tax and Budget Bulletin*, September 2003, 2 pages.
- J. Gayek, L. Long, K. Bell, R. Hsu, and R. Larson, "Software Cost and Productivity Model," Ground Systems Architecture Workshop, March 2004, 50 charts.
- General Accounting Office, *Defense Acquisitions: Improved Management Practices Could Help Minimize Cost Growth in Navy,* GAO-05-183, February 2005, 80+iii pages.
- General Accounting Office, *Defense Acquisitions: Assessments of Major Weapons Programs*, GAO-03-476, May 2003, 74+iv pages.
- General Accounting Office, *NASA Program Costs*, GAO Report NSIAD 93-97, December 1992, 21 pages.
- J. Hihn and H. Habib-agahi, "Flight software cost growth: analysis and recommendations," NASA Cost Estimating Symposium, March 2000, 15 charts, www.ipao.larc.nasa.gov/symposium/SW-Hihn.pdf
- D. McNicol, "OSD CAIG Cost Growth Study," National Defense Industrial Association, May 2001, 20 charts, www.dtic.mil/ndia/2001sbac/mcnicols.pdf
- Office of the Undersecretary of Defense for Acquisition, Technology, and Logistics, *Report of the Defense Science Board/Air Force Scientific Advisory Board Joint Task Force (Young Commission) on Acquisition of National Security Space Programs*, May 2003, 74+vi pages.
- Pacific Northwest National Laboratory, *Evidence of Cost Growth Under Cost-Plus and Fixed-Price Contracting*, PNNL-11984, September 1998, 27+vi pages.
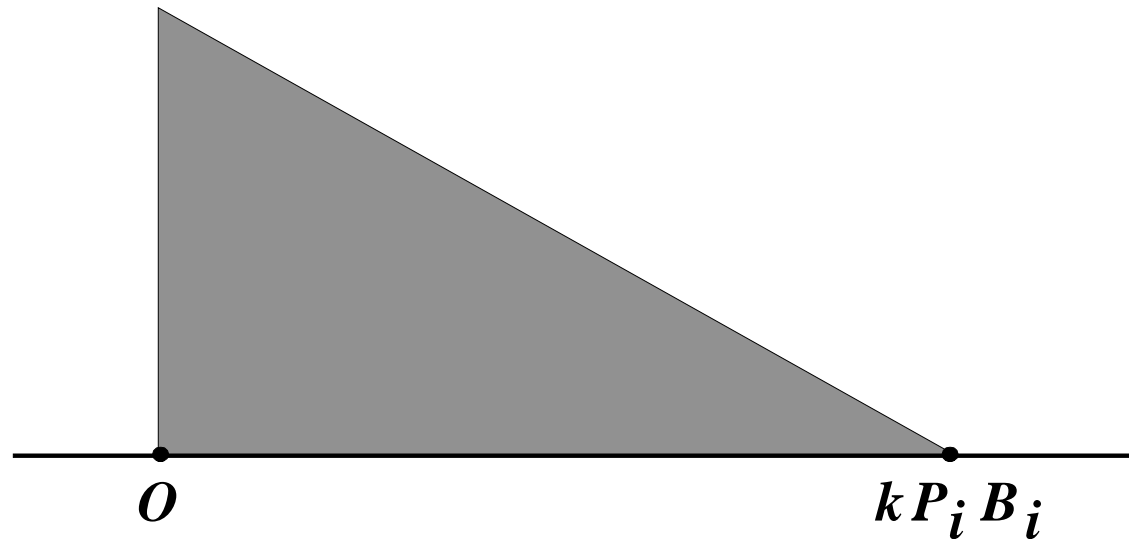
B_i = Point Estimate Cost of Element i

P_i = Fraction of New Technology in Element i

k = Cost-Growth Sensitivity Factor

- **Costs of WBS Elements i and j are**

$$C_i = B_i + E_i + R_i, \quad C_j = B_j + E_j + R_j$$

- **Inter-Element Correlations between WBS Elements i and j are** $\rho_{ij} = Corr\left(C_i, C_j\right)$

- **Correlations between Following Pairs of Random Variables are All Zero:**

$Corr\left(B_i, B_j\right), \; Corr\left(B_i, E_j\right), \; Corr\left(B_j, E_i\right), \; Corr\left(B_i, R_j\right),$
$Corr\left(B_j, R_i\right), \; Corr\left(E_i, E_j\right), \; Corr\left(E_i, R_j\right), \; Corr\left(E_j, R_j\right)$

- **But** $Corr\left(R_i, R_j\right) = 1$

- $Var(R_i) = \dfrac{(kP_iB_i)^2}{18}, \quad Var(R_j) = \dfrac{(kP_jB_j)^2}{18}$

$$Var(C_i) = U_i^2 + \dfrac{(kP_iB_i)^2}{18}, \quad Var(C_j) = U_j^2 + \dfrac{(kP_jB_j)^2}{18}$$

$$Cov(C_i, C_j) = Cov(R_i, R_j)$$

- **From the Assumption that** $Corr(R_i, R_j) = 1$,

  it Follows that …

$$\rho_{ij} = Corr\left(C_i, C_j\right) = \frac{Cov\left(C_i, C_j\right)}{\sqrt{Var(C_i)\ Var\left(C_j\right)}} = \frac{Cov\left(R_i, R_j\right)}{\sqrt{Var(C_i)\ Var\left(C_j\right)}}$$

$$= \frac{Corr\left(R_i, R_j\right)\sqrt{Var(R_i)\ Var\left(R_j\right)}}{\sqrt{Var(C_i)\ Var\left(C_j\right)}}$$

$$= \frac{\sqrt{\dfrac{\left(kP_iB_i\right)^2\left(kP_jB_j\right)^2}{18^2}}}{\sqrt{U_i^2 + \dfrac{\left(kP_iB_i\right)^2}{18}}\ \sqrt{U_j^2 + \dfrac{\left(kP_jB_j\right)^2}{18}}}$$

$$= \frac{k^2 P_i P_j B_i B_j}{\sqrt{18U_i^2 + k^2 P_i^2 B_i^2}\ \sqrt{18U_j^2 + k^2 P_j^2 B_j^2}}$$

$$\mu_i = \text{Mean of Element } i \text{ Cost} = B_i + \frac{kP_iB_i}{3}$$

$$\sigma_i^2 = \text{Variance of Element } i \text{ Cost} = U_i^2 + \frac{(kP_iB_i)^2}{18}$$

$$\rho_{ii} = \text{Correlation Between Costs of Elements } i, j$$

$$= \frac{k^2 P_i P_j B_i B_j}{\sqrt{18\ U_i^2 + k^2 P_i^2 B_i^2}\ \sqrt{18\ U_j^2 + k^2 P_j^2 B_j^2}}$$

- **We Started with …**
  - … WBS-Element Cost Probability Distributions
  - … A Model of How They Could be Correlated
- **We then Derived Inter-WBS-Element Correlations**
- **Furthermore, These Correlations are Guaranteed to Form a <u>Mathematically</u> Valid* Correlation Structure, because They Have Been Derived from Mathematically Valid Statistical Relationships**
- **The Formulas will be Slightly Different if the Correlations $Corr\left(R_i, R_j\right)$ are Less than 1.0**
- **The Correlations are Now Ready to be Input into Software that Computes the Total System Cost Probability Distribution**

* The correlation matrix will have no negative eigenvalues.

Contents

- **Cost-Risk Analysis**
- **Limits of Engineering Judgment**
- **The Trouble with Inputs**
- **A Quantification Model**
- **Modeling Inter-Element Correlation**
- **Examples**
- **Summary**

Example: System Y (Basic) Program-Descriptive Inputs

WBS Element	Point Estimate ($M)	Standard Error (as % of Point Estimate)	Percentage New (%)
1. Satellite	200	30	50
2. Launch	80	10	0
3. Ground	100	20	30
4. Data Distribution	300	40	20

Example: System Y (Basic) Correlation Matrix for $k = 1.0$

WBS Element	1	2	3	4
1	1.00	0.00	0.12	0.04
2		1.00	0.00	0.00
3			1.00	0.04
4				1.00

Note: k = Cost-Growth Sensitivity Factor, a Measure of the Extent to which the "Newness" Percentage is Expected to Impact Cost Growth

Example: System Y (Basic) Correlation Matrix for $k = 2.0$

WBS Element	1	2	3	4
1	1.00	0.00	0.36	0.14
2		1.00	0.00	0.00
3			1.00	0.13
4				1.00

Note: k **= Cost-Growth Sensitivity Factor, a Measure of the Extent to which the "Newness" Percentage is Expected to Impact Cost Growth**

Example: System Y (Basic) Correlation Matrix for $k = 3.0$

WBS Element	1	2	3	4
1	1.00	0.00	0.55	0.25
2		1.00	0.00	0.00
3			1.00	0.24
4				1.00

Note: k = **Cost-Growth Sensitivity Factor, a Measure of the Extent to which the "Newness" Percentage is Expected to Impact Cost Growth**

Example: System Y (Basic) Correlation Matrix for $k = 4.0$

WBS Element	1	2	3	4
1	1.00	0.00	0.69	0.36
2		1.00	0.00	0.00
3			1.00	0.35
4				1.00

Note: k = Cost-Growth Sensitivity Factor, a Measure of the Extent to which the "Newness" Percentage is Expected to Impact Cost Growth

Example: System Z (New Technology) Program-Descriptive Inputs

WBS Element	Point Estimate ($M)	Standard Error (as % of Point Estimate)	Percentage New (%)
1. Satellite	200	30	90
2. Launch	80	10	10
3. Ground	100	20	60
4. Distribution	300	40	70

Example: System Z (New Technology) Correlation Matrix for $k = 1.0$

WBS Element	1	2	3	4
1	1.00	0.13	0.33	0.22
2		1.00	0.13	0.09
3			1.00	0.22
4				1.00

Note: k = **Cost-Growth Sensitivity Factor, a Measure of the Extent to which the "Newness" Percentage is Expected to Impact Cost Growth**

Example: System Z (New Technology) Correlation Matrix for $k = 2.0$

WBS Element	1	2	3	4
1	1.00	0.35	0.67	0.52
2		1.00	0.35	0.27
3			1.00	0.52
4				1.00

Note: k = Cost-Growth Sensitivity Factor, a Measure of the Extent to which the "Newness" Percentage is Expected to Impact Cost Growth

Example: System Z (New Technology) Correlation Matrix for $k = 3.0$

WBS Element	1	2	3	4
1	1.00	0.52	0.82	0.70
2		1.00	0.52	0.45
3			1.00	0.70
4				1.00

Note: k **= Cost-Growth Sensitivity Factor, a Measure of the Extent to which the "Newness" Percentage is Expected to Impact Cost Growth**

WBS Element	1	2	3	4
1	1.00	0.65	0.89	0.81
2		1.00	0.65	0.59
3			1.00	0.81
4				1.00

Note: k = **Cost-Growth Sensitivity Factor, a Measure of the Extent to which the "Newness" Percentage is Expected to Impact Cost Growth**

Contents

- **Cost-Risk Analysis**
- **Limits of Engineering Judgment**
- **The Trouble with Inputs**
- **A Quantification Model**
- **Modeling Inter-Element Correlation**
- **Examples**
- **Summary**

- **Costs are Random Variables, Not Deterministic Numbers, with Cost Risk (Uncertainty) Due to**
  - **Estimating Error**
  - **Probable Technical Difficulties**
- **Cost Risks of Different Program Elements are Correlated, with Inter-Element Correlations Difficult to Estimate (i.e., any Numerical Values are Difficult to Justify)**
- **Coherent Theory Presented Here to …**
  - **… Replace Challenge of Estimating Correlations by the Need to Estimate Percent of Required Technology that is New**
  - **… Derive "Cost-Growth Sensitivity Factor" from Historical Experience**
  - **… Provide Theoretical Justification for a Set of Numerical Values of Inter-Element Correlations**